

ABSTRAK

Muhammad Fakhri Abdillah, 2023. *Implementation of Random Forest Algorithm for Predicting Imbalanced Multiclass Data by Applying Stratified K-Fold and Leave One Out Cross Validation.* **SKRIPSI.** Gorontalo. Program Studi Statistika. Jurusan Matematika. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Negeri Gorontalo.

Pembimbing : (1) **Dra. Lailany Yahya, M.Si**, (2) **Muhammad Rezky F. Payu, M.Si**

Data *multiclass* tidak seimbang merupakan masalah yang sering dihadapi dalam metode pembelajaran *supervised learning* khususnya klasifikasi. Ketidakseimbangan kelas mengakibatkan kelas minoritas kurang terwakili sehingga prediksi yang dilakukan menjadi kurang baik. Teknik *cross validation* yang digunakan terhadap masalah tersebut adalah *Stratified K-Fold* dan *Leave One Out*. Kedua teknik ini akan diterapkan pada metode klasifikasi *Random Forest* untuk melihat hasil prediksi terbaik diantara keduanya yang dilihat dari sisi akurasi, *recall* dan presisi. Hasil penelitian menunjukkan bahwa pengklasifikasian *Random Forest* dengan menerapkan *Leave One Out* menghasilkan nilai yang lebih tinggi dibandingkan dengan *Stratified K-Fold*. Hal ini menandakan bahwa *Leave One Out* baik untuk diterapkan terhadap data *multiclass* tidak seimbang.

Kata Kunci: *Data multiclass tidak seimbang, Random Forest, Cross Validation, Stratified K-Fold, Leave One Out*

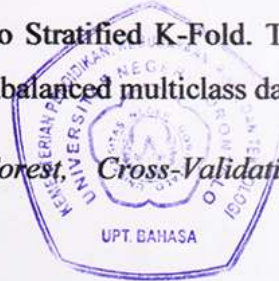
ABSTRACT

Muhammad Fakhri Abdillah, 2023. *Implementation of Random Forest Algorithm for Predicting Imbalanced Multiclass Data by Applying Stratified K-Fold and Leave One Out Cross-Validation.* **UNDERGRADUATE THESIS.** Gorontalo. Study Program of Statistics. Department of Mathematics. Faculty of Mathematics and Natural Sciences. Universitas Negeri Gorontalo.

Supervisors : (1) **Dra. Lailany Yahya, M.Si**, (2) **Muhammad Rezky F. Payu, M.Si**

Imbalanced multiclass data are a problem that is often encountered in supervised learning methods, especially classification. The imbalanced class results in the under-representation of the minority class, so the predictions made are less good. The cross-validation techniques used for this problem were Stratified K-Fold and Leave One Out. These two techniques were applied to the Random Forest classification method to determine the best prediction results between the two in terms of accuracy, recall, and precision. The research finding denoted that the classification of Random Forest using Leave One Out obtained a higher score compared to Stratified K-Fold. This result signified that Leave One Out was better to apply to imbalanced multiclass data.

Keywords: *Imbalanced Multiclass Data, Random Forest, Cross-Validation, Stratified K-Fold, Leave One Out*



LEMBAR PENGESAHAN

Skripsi yang berjudul "**Implementasi Algoritma *Random Forest* untuk Memprediksi Data *Multiclass* Tidak Seimbang dengan Menerapkan *Stratified K-Fold* dan *Leave One Out Cross Validation*"**

Oleh

MUHAMMAD FAKHRI ABDILLAH
NIM. 413419014

Program Studi Statistika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Telah dipertahankan di depan pembimbing dan dewan penguji

Hari, tanggal : Selasa, 13 Juni 2023

Waktu : 13.00-14.00 WITA

Tempat : Ruang Sidang Jurusan Matematika

A. Pembimbing

1. **Dra. Lailany Yahya, M.Si**

NIP. 196812191994032001

Penguji Utama

Tanda Tangan



2. **Muhammad Rezky F. Payu, M.Si**

NIP. 198912152018031003

Anggota



B. Dewan Penguji

3. **Nurwan, S.Pd., M.Si**

NIP. 198105102006041002

Anggota

Tanda Tangan



4. **Salmun K. Nasib, S.Pd., M.Si**

NIP. 198903302019032018

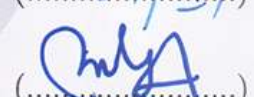
Anggota



5. **Nisky Imansyah Yahya, S.Pd., M.Si**

NIP. 199107302020121008

Anggota



Mengetahui,

Dekan Fakultas Matematika dan IPA



Prof. Dr. Astin Lukum, M.Si

NIP.196303271988032002